

WEATHER PREDICTION ANALYSIS USING DATA MINING

S.LATHAMrs. V. SHOBANA

II-M.Sc.Computer Science, Assistant Professor,

Department of computer science, Department of computer science,

Coimbatore.Coimbatore.

Abstract—In present situation weather forecasting and analysis has become a stimulating problem around the world from the last century. The reason behind are the two main factors: Firstly, it is valuable for many human activities like agriculture sector, tourism and natural disaster prevention. Secondly, due to various technological advances like the growth of computational power and ongoing developments in measuring systems. All over the world, major challenges faced by meteorologist are the precision of weather analysis and its prediction. Data mining matters is to provide correct knowledge in the form of useful rules, techniques, visual graphs and models for the weather parameters over the datasets. This information can be utilized to support many important areas which are affected by climate change includes Agriculture, Water Resources, Vegetation and Tourism. Study shows that human society is affected in different ways by weather affect. For example, water resources are the main sources of irrigation in production of agriculture crops and the amount of rain is one of them that affect the crops abruptly due to climate change.

Keywords—

Weatherprediction,Datamining,Decision tree,C4.5 algorithm.

I. INTRODUCTION

Weather prediction has been a stimulating problem in meteorological department since years. Even after the technological and scientific progress, the accuracy in prediction of weather has never been sufficient. Even in current date this domain remains as a research topic in which scientists and mathematicians are working to produce a model or an algorithm that will accurately predict weather [1].

A . DATA MINING TECHNIQUES

Data mining objects is to provide accurate knowledge in the form of useful rules, events, visual graphs and models for the weather parameters over the datasets. This knowledge can be used to care the decision-making for various sectors. The areas for data analysis are those which include weather differences that affect our daily runtime changes in min and max temperature, humidity level, rainfall chances and speed of wind. This information can be utilized to support many important areas which are affected by climate change embraces Agriculture, Water Resources, Vegetation and Tourism [2].Data mining is the process of applying these methods with the intention of uncovering hidden patterns in large data sets [4]. Data mining is the application of specific algorithms for extracting patterns from the huge data.

B. DECISION TREE

A decision tree is a decision boost device that uses a tree-like chart or perfect of decisions and their conceivable results, counting chance occasion results, asset expenses, and utility. It is one method to show a calculation. A Decision Tree is a torrent outline like tree structure. Each node indicates a test on a property. Every branch expresses to a result of the test. Leaf nodes speak to class appropriation. The choice tree structure gives an express plan of "assuming then" guidelines making the results simple to translate [5]. Tree-shaped structure that signifies sets of decisions. These results make rules for the organization of a dataset. Exact decision tree methods include Classification and Regression Trees (CART) and ID3 [6].

II. LITERATURE REVIEW

Many scholars have made efforts to tool different mining techniques in the areas of meteorological data created on weather data analysis and prediction. Meteorology data mining has been successfully employed in the field of developing important forecasting applications [2]. Literature survey provides the mandatory knowledge about the project and its background. It also helps in resulting the best practices in project development. Literature survey also helps in understanding the risk and viability of the project. Weather historical data is composed from Indian Meteorological Department (IMD) Pune. From the collected weather data attributes which are most relevant to rainfall prediction are chosen. Data pre-processing and data alteration on raw weather data set is performed, so that it shall be conceivable to work on Bayesian. The model is trained using the training data set and has been verified for accuracy on accessible test data. The meteorological centers use high performance computing and supercomputing power to run weather prediction model. Author proposed and implemented data intensive model using

data mining technique. The model works with good accuracy and takes moderate compute resources to predict the rainfall [3]. The different arrangement methods namely Decision Trees, Rule-based Methods, Neural Networks, Naïve Bayes, Bayesian Belief Networks, and Support Vector Machines. This paper references different personalities like decision-tree algorithms like ID3 (Iterative dichotomiser3), C4.5 (successor of ID3), CART (Classification and Regression Tree). Decision tree is the considered as the powerful solution to the classification problems and it is applied in many real world applications [5]. Many data mining techniques are used for weather forecasting in the present scenario with various level of accuracy.

III. MATERIALS AND METHODS

Weather is one of the maximum influential factors in our daily life, to an amount that it may affect the economy of a country that depends on occupation like agriculture. Therefore as a counter amount to reduce the damage caused by the uncertainty in weather behavior, there should be efficient ways to predict weather. Usually two main methods are used for weather forecasting, one includes usage of large amount of data to gain information about future weather and the other includes construction of equations that will help forecast weather by identifying different parameters and substituting the values to obtain desired result. Recently investigators have started highlighting the effectiveness of data algorithms in predicting weather. Various data mining techniques and gave a performance comparison between algorithms like C4.5, CART. This stage consisted of analyzing the given data-set with different algorithms like Naïve Bayes and C4.5 (J48) algorithm and then choosing the better one for further predictions. Then the dataset was split into training set for making the machine learn and the testing dataset along with cross validation. Then the patterns were recorded to make further predictions [1].

A.C4.5 DECISION TREE

Unlike Naïve Bayes, the C4.5 is arrangement algorithm used to generate decision tree for the given dataset. It is based on the information entropy concept. Creation of the decision tree is done by selecting the best possible attribute that will be able to split set of samples in most actual manner. The attribute having the maximum entropy difference or regularized information gain is selected as the unbearable criteria for that specific node. Similar fashion is followed and nodes are added to the decision tree. Each penultimate node transmits the last attribute or many attributes for making the final decision of the tricky.

Algorithm J48:

```
INPUT
D // Training data
OUTPUT
T // Decision tree
DTBUILD (*D)
{
T = Null;
T = Create root node and label with splitting
attribute;
T = Add arc to root node for each split predicate
and label;
For each arc do
D = Database created by applying splitting
predicate to D;
If stopping point reached for this path, then T'=
Create leaf node and label with appropriate class;
Else
T' = DTBUILD (D);
```

After the performance comparison, the J48 algorithm was chosen for further implementation which involved study of the legacy data about the weather [1].

B.NAIVE BAYES

They can forecast class association prospects such as the probability that a given tuple fits to a

particular class. The Bayesian Classifier is capable of scheming the most probable output depending on the input. A Naive Bayes classifier receives that the attendance (or nonattendance) of a particular feature of a class is unconnected to the attendance (or nonattendance) of any other feature, given the class variable [3].

IV. RELATED WORK

Meteorological conditions forecast has always been a fascinating field of investigation as changes in climatic conditions straight impact the population has classified the approach of weather forecast in two category, Experimental approach and Numerical approach. Experimental approach collects present weather conditions through ground comments i.e. observations from ships, satellites etc . [6]. Yuko Tachibana and Mikihiro Ohnari grants Model of Hourly Water Consumption in Water Cleansing Plant through Definite Method because the need of forecast of hourly water consumption while providing water to consumers. Water consumption is shown in waveform and is inclined of the week weather and temperature they look like each other. Water consumption on breaks and other days are not same. By examining the water consumption in metropolitan cities and by applying data mining techniques constructed precise forecast for year. Charles a. Doswell et.al, examines the method for flood forecasting. It is the result of high rainfall rates. This contains water air in the air and also be contingent on water precipitation [7].

V. PROPOSED WORK

This systems information is sent to meteorological centers where the data are collected, analyzed, and made into a variety of charts, maps, graphs and data sets. Modern high-speed computers transmission the many thousands of comments onto surface and upper-air maps. A final data conventional is collected for an analysis. Weather forecasting has been one of the” most systematically and

technologically inspiring problems around the world in the last century”. “This is due mainly to two factors: first, it’s used for many human actions and secondly, due to the opportunism twisted by the various technological advances that are directly related to this concrete research field, like the evolution of computation and the improvement in measurement systems [5]. Future prediction with the effects of the temperature along with defensive measures. To predict the weather conditions in the proposed system we used the data mining algorithms [7]. The future framework which is based on Arbitrary Decision Tree stretches above 80% accuracy for forecasting weather of different cities of dissimilar countries, which is very respectable accuracy in machine learning classification [8].

VI. CONCLUSION AND FUTURE ENHANCEMENTS

For the current application of data mining in weather prediction domain, the analysis of Naïve Bayes and C4.5 Decision Tree algorithm was done simultaneously with dataset containing weather data collected over a period of 2 years. It was found that the presentation of C4.5 (J48) decision tree procedure was far better than that of Naïve Bayes. Previous research has highlighted that performance of C4.5 algorithm improves when the dataset used for application is quite large whereas the performance of Naïve Bayes is comparatively poor. The accuracy for C4.5 was 88.2% with respect to classifying the instances correctly. On the other hand, Naïve Bayes showed a poor performance of 54.8% while classifying the instances.

REFERENCES

1. Analysis of Data Mining Techniques for Weather Prediction, Fahad Sheikh¹, S. Karthick, D. Malathi, J. S. Sudarsan and C. Arun, Department of Software Engineering, SRM University, Kattankulathur, Chennai - 603203, Tamil Nadu, India; fahadsheikh1709@gmail.com, karthik.sa@srmuniv.ac.in, arun.c@ktr.srmuniv.ac.in
2. Department of Computer Science and Engineering, SRM University, Kattankulathur, Chennai - 603203, Tamil Nadu, India; malathi.d@ktr.srmuniv.ac.in
3. Department of Civil Engineering, SRM University, Kattankulathur, Chennai - 603203, Tamil Nadu, India; sudarsanjss@yahoo.com, Indian Journal of Science and Technology, Vol 9(38), DOI: 10.17485/ijst/2016/v9i38/101962, October 2016.
2. Application of Data Mining Techniques in Weather Data Analysis, M Ramzan Talib, Toseef Ullah, M Umer Sarwar, M Kashif Hanif and Nafees Ayub, Department of Computer Science, Government College University, Faisalabad, Pakistan, IJCSNS International Journal of Computer Science and Network Security, VOL.17 No.6, June 2017.
3. Weather Forecasting Using Data Mining, [1] Siddhant Revankar, [2] Shaba Desai [1] Department of Information Technology, Padre Concilio College of Engineering Verna, Goa, Goa University International Journal of Engineering Research in Computer Science and Engineering (IJERCSE) Vol 5, Issue 2, February 2018.
4. Decision Tree for the Weather Forecasting, Rajesh Kumar, Ph.D Asst. Prof., Dept. of ECS Dronacharya College

- of Engineering, Gurgaon, India. International Journal of Computer Applications (0975 – 8887) Volume 76 No.2, August 2013
5. Weather Prediction Based on Decision Tree Algorithm Using Data Mining Techniques, Siddharth S. Bhatkande¹, Roopa G. Hubballi² Department of CSE, KLE Dr. MSS College of Engg& Tech. Belgaum India ¹ Professor, Computer Science & Engg, KLE DR M S Sheshgiri College of Engg& Tech., Belgaum², International Journal of Advanced Research in Computer and Communication Engineering Vol. 5, Issue 5, May 2016.
 6. Data Mining Techniques Using To Weather Prediction, P. Kalaiyarasi [1], Mrs. A. Kalaiselvi [2] Research Scholar [1], Assistant Professor [2] Navarasam Arts and Science College for Women, Arachaur India, International Journal of Computer Science Trends and Technology (IJCST) – Volume 6 Issue 3, May - June 2018.
 7. Weather prediction using the data mining Techniques, Aishwarya Dhore¹, Anagha Byakude², Bhagyashri Sonar³, Mansi Waste⁴; ^{1,2,3,4} DYP Pimpri College of Engineering, Pune, Maharashtra, India, International Research Journal of Engineering and Technology (IRJET) Volume: 04 Issue: 05 | May - 2017.
 8. Arbitrary Decision Tree for Weather Prediction, Nalanda B Dudde¹, Dr. S. S. Apte² ¹ Walchand Institute of Technology, Solapur, India ² HOD Walchand Institute of Technology, Solapur, India, International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064 Index Copernicus Value (2013): 6.14 | Impact Factor (2014): 5.611.